

Techniques, Applications and Challenges of Opinion Mining

Kazi Mostafizur Rahman* and Aditya Khamparia**

ABSTRACT

Opinion targets and words from online feedback is an essential and challenging work. It is a Natural language processing and information extraction task to identifies in the way to get about positive, negative or neutral comments and quotes in the text. There are various test result listed in the table researched in Opinion Mining and few key challenges listed below. This survey gives an overview idea about the future research directions in the field of Opinion mining which will help us to enhance the research for future in recommendation system.

Keywords: Data Mining, Opinion Mining, Sentiment Analysis, Knowledge discovery in database, Sentiment classification

1. INTRODUCTION

The exponential growth of internet usage has created a new platform where people can freely communicate and exchange ideas and opinions. [3] The Web has important, large and unstructured information about opinion. User's opinion can be important when it is about to make any decision or selection from different various products. That includes important resource- e.g from past experiences it is helpful to spend money on any product or service, which gives satisfaction to customers. Now "Social Media" really

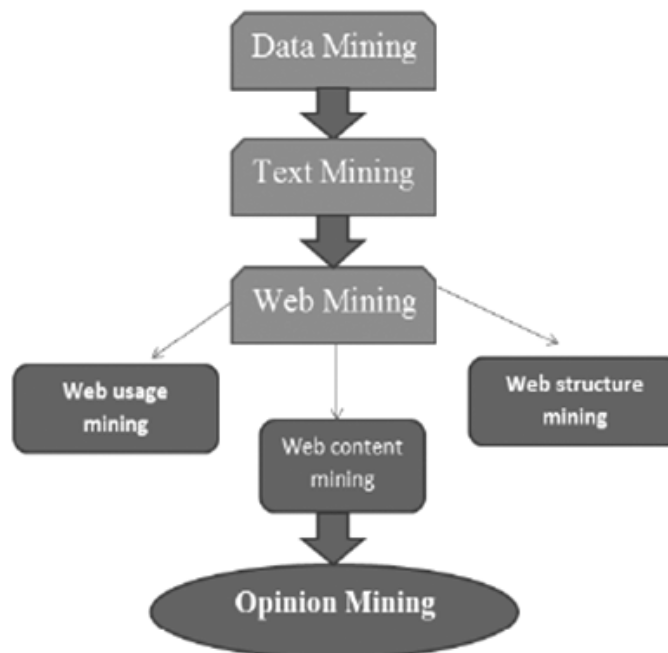


Figure 1: Data Mining Hierarchy

* Department of Computer Science and Engineering Lovely Professional University Jalandar, Punjab, India, Email: kaziraja@gmail.com

** Department of Computer Science and Engineering Lovely Professional University Jalandar, Punjab, India, Email: aditya.khamparia88@gmail.com

helps users to generate and share ideas among them connected through internet. Websites like Blogs, Social Networks, e-commerce etc helps customers to know about valuable information. This sort of details is not structured and because it is created for users, it is not something which is generated “machine processable” [13][14]

To save the large amount of data there are many various techniques and models are required to get the information from unstructured pattern and knowledge. These techniques are used to convert the data into valuable information to analysis the business, detect frauds and get an overview idea about customer view or interest etc. These kinds of techniques are said to be Knowledge Discovery (KD) in Data Mining(DM). [1][2]Opinion Mining is the area to calculate the opinion in the text form which is easy understand for user. Also get an idea about the customer likes and dislikes or any other view (neutral) which is extraction of information from unstructured data. In online sites, web content mining, involves Natural language Processing (NLP) and feedback analysis is done to get knowledge about popular product feedback by users. [1]

Consumers wants to know product for which they read about product and its reviews and try to make decision to buy any product. It is observed that feedbacks are getting necessary to analysis and these feedbacks are benefited for the companies and the consumers for sell or buy any product. [28] The count of reviews can be in large number for a demanding item that can be difficult for users to read all of them and come to an conclusion to buy the item. Sometimes it also makes it challenging for the industry to keep track of all feedbacks mentioned by users. To an extent it is possible but time consuming. For example, manufacturing companies refer to information in a format that is understandable form [29]. Due to which opinion mining comes to existence and can be expressed in format of image or text, video or audio which is use to use for users.

1.1. Main Challenges/difficulties in OM

Customer feedbacks or status can be in many language (English,Hindi,Urdu,Arabicetc) so it is challenging to understand each language according to it. Grammar- Noun words are also be said as featured words. but Verbs and Adjective can also be used as feature words which are challenging to judge.[31] When any user like -RAM gave a feedback on earphone, “The sound quality is very good” and RAMAN’s feedback is “The sound is awesome” here RAM and RAMAN is sharing the quality of the product but in another manner. The Similar/opposite words are also difficult task to group [35]. For example “Camera size of mobile phone is small”. Here adjective “small” used in +ve manner but if user said that “The battery time is also small” now the small represent -ve to battery phone. To identify such adjective words which define different meaning in same situation is a difficult to handle it.. [41]

Users are allowed to feedback in any format, they are free to write in any manner. Users can use symbols, abbreviation, capital letter or small letter, shortcut and regional language in feedbacks e.g camra as camera, pix as pictures,fi9 as fine, guuds as goods etc. It is difficult to handle such kind of data, requires to mine point oif view of users. [37]Many users has different point of view using same kind words in their feedback which is challenging to sort out the +ve,-ve or neutral meaning while mining information. [37]

1.2. Why this Opinion mining service is conducted?

To identify any spam and fake reviews, the comparison of few reviews and detection of outliers and the reputation of the reviewer. [38] The combination of opinion with behavior might give a different data of opinion expressed. [39] For example - Can be misinterpret with the word or sentence and the strikers in comment or feedback. That is - I lost my phone (smiling striker), it is difficult to analysis that the user is happy as he might get a new phone or sad as he lost it.

2. TECHNIQUES USED IN OPINION MINING

Database is used for decision making from the available data or extract information. Different database like relational, object oriented, transactional etc consist on the complex dataset. The huge number of data in databases has made this necessary to make technologies to get the idea and information [4]. DM techniques use to get an idea and information like supervised, unsupervised, generalization, Case Based Reasoning (CBR), fuzzy logic, genetic algorithm association rule mining, data visualization, neural network, Bayesian networks, genetic algorithm, decision tree, multi agent systems, CRISP-DM model, churn prediction and many more.[1]

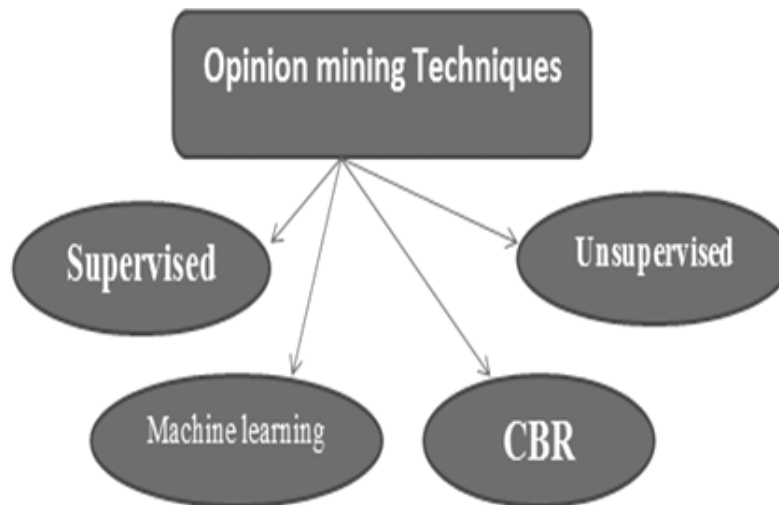


Figure 2: Techniques of Opinion Mining

2.1. Supervised Machine Learning

The supervised machine learning is that one in all the techniques Classification is used data processing technique [1] [13]. Polarity classification is the basic task of opinion mining. It basically used to positive or negative polls for example- in feedback such as “thumbs up” versus “thumbs down” or “like” versus “dislike”. Polarity classifications also identify pro and con expressions in online feedbacks and helps to make decision to buy any particular product. [2]

2.2. Unsupervised Machine Learning

Unlike Supervised Machine Learning, this learning has no fix parameter to judge an output related to input.[4] It is process to learn by observation. Clustering is the technique utilized in unsupervised learning. The process of gathering objects of similar data characteristics into a similar group is called clustering. In the group on clustering objects in one cluster are dissimilar to the objects in other clusters by this the observation is done [1] [3].

2.3. Case Based Reasoning

CBR is a prominent AI classification technique that is to get the conclusion of a new given problem which is depends on old similar kind of problems. CBR is a technique of computer reasoning which solves cases that is close to real time scenario which is called problem solving techniques where knowledge is represent as past cases in stored and that does not rely in any defined rules. The old case’s solutions are stored in CBR repository called Knowledge base. CBR uses the idea from the Knowledge base to again use the solution of old similar issue. When required to make any changes or solve any problem observe the old similar case from knowledge base. [1][4]

3. APPLICATION DOMAIN

Here we are discussing about few opinion mining application domain.

3.1. Shopping (E-Commerce)

Online Shopping is important for people as it deliver package to their door steps because it is easy to get an overall idea about any product by the number of feedbacks mentioned by customers. In this kind of site users go through the feedback before they buy any product. [5] Shopping sites Flipkart, Amazon helps to compare products with all desired description and feedback of customers, information is displayed to the user in a “Graphical User Interface” for easy understanding about the quality/features and services of product. This is what really helps users to get their selection to product want to buy.

3.2. Entertainment

For Movie or show or TV programs can easily go through public feedback about any current release or any kind of popular program or movie. Internet movie database (IMDB) which helps the users online to view feedback for movies and other programs. It also helps users to understand the movie or program who are not sure it. [5].

3.3. Business

Companies are now asking for feedbacks on their portal which makes savings on marketing budget. Now companies can be able to analysis the data they from the given feedback online in their website. Recommendation to friends and family about products or services to each other which helps to gain an clear picture about the products or services before buy it. [5] [15]

3.4. R&D (Research and Development)

In any online shopping portal items feedback are used by industry to improve the quality of the products. Online sites can also ask the customers to provide their own design views to make modification or create new products for customers [5]. For example “Play Store” in which it shows all the applications users has given all sort of feedback/point of view.

3.5. Politics

Political people can extract the views on public feedback. During Elections the participated candidates can have an overall knowledge about public opinion which helps to gain an idea about the progress (weakness and strength)[5].

3.6. Academics

In any online course, Student’s opinion can be used to manage their academic. Student’s feedback may help the institute to understand and analysis and make better improvement for student’s future benefits. 5]. Institution will be aware of their benefits and problems through their opinion in the form on feedback and comments.

3.7. Health

User’s shares treatment which helps to take precaution while they suffer from headache, fever, high blood pressure, Diabetes and many more which is also called “Home Remedies”. Users also share the exercise steps for body building, blood pressure, back pain, migraine etc which helps all to get rid of problems in their life. While doing so users also gives feedback and comment from which is it understandable that particular remedies is working or not to others.

3.8. Transportation

Transportation can be any movable items like animals, people, and goods from one place to another as per the requirement. There are many transport modes for example road, air, rail, water, cable, space and pipeline. Transport is necessary to trade between persons, which is essential for the development of civilizations. Transportation like tour and travel, logistic, cab service which is very helpful to people to choose best services mentioned in the feedback and comments from past experience shared by the users. For example from Tour and Travel that is holiday plans it helpful to get an idea about the location and hotels and services provided to them

Figures and Table

| 1 | Author | Dataset | Technique | Result (Accuracy) |
|----|-----------------------------|----------------------|---------------------------------------|--------------------|
| 2 | Pang and Lee (2004) | Movie Review | Naive Bayes support vector machines | 86.40% |
| 3 | Gamon (2005) | Car reviews | Naive Bayes | 86% |
| 4 | Hu and Liu (2005) | Amazon Cnn.Net | Opinion word extraction | 73% |
| 5 | Bai . (2005) | Movie review | two-stage Markov Blanket Classifier | 87.50% |
| 6 | Popescu and Etzioni (2005) | Amazon | relaxation labeling clustering | 86% |
| 7 | Kennedy and Inkpen (2006) | Online Site | support vector machines, termcounting | 86.20% |
| 8 | Korrig ,Brill(2006) | Movie review | Hybrid | 91% |
| 9 | Godbole et al. (2007) | blog posts | Lexical approach | 95.70% |
| 10 | Songho tan (2008) | Chnsenticorp | SVM | 90% |
| 11 | Zhou and Chaovalit (2008) | Movie Review | ontology-supported polarity mining | 72.20% |
| 12 | Melville (2009) | Blogs | Bayesian classification | 91.21% |
| 13 | Rudy(2009) | movie review | ID3,SVM,Hybrid | 89% |
| 14 | QingliangMiao (2009) | Amazon reviews | Lexical resource | 87.60% |
| 15 | Yulan He (2010) | movie review | Sentiment lexicon | 74% |
| 16 | Zhu Jian (2010) | movie review | Back propagation | 86% |
| 17 | KaiquanXu(2011) | Amazon Reviews | Multiclass SVM | 61% |
| 18 | Ziqiong (2011) | Cantonese reviews | Naive bayes,SVM | 93% |
| 19 | Ayesha Rashid (2013) | Indian Hotel Reviews | NLP and Bayesian | 96.06% |
| 20 | S. Vasantharaj et al (2015) | China cars reviews | Word Kernel Path Kernel N-gram Ker | 80.93% |

4. CONCLUSION

In this survey paper, we have researched from the year “ 2004-2015” which is mentioned in Table and its graphical representation with few techniques and dataset in Opinion mining.. OM is a prominent area of data extraction to get knowledge from large amount of data from users. Although there are many drawbacks exist in sentiment classification, many difficulties and gap like grammar mismatch, writing technique, regional language, fake detection feedbacks etc. For the future research areas of opinion mining scope is researched which is sentimental analysis is in trend, that is feedbacks of “twitter” comments or statements, status or comments on pictures, videos, status etc and even on other online shopping sites product reviews. An overview of multiple opinion mining which are Classification, Clustering and Case based reasoning (CBR) techniques is examined to provide an clear picture to future research which helps to get users recommend to make a selection that can be also called “Recommendation System”

REFERENCES

- [1] Ayesha Rashid, Naveed Anwer, Dr. Muddaser Iqbal, Dr. Muhammad Sher, “Areas, Techniques and Challenges of Opinion Mining”, *IJCSI International Journal of Computer Science Issues*, in 2013
- [2] S. Chandra Kala and C. Sindhu, “Opinion Mining and Sentiment Classification”, *IJSC Vol 3* in 2012
- [3] Bing Liu, “Sentiment Analysis and Opinion Mining” Morgan & Claypool Publishers in April 22, 2012
- [4] S. Vasantharaj, “A Survey on Sentiment Analysis Applied in Opinion Mining”, *Journal of Network Communications and Emerging Technologies (JNCET)*, March (2015)
- [5] Dipali V. Talele “The Art of Opinion Mining and Its Application Domains”, *International Conference in Recent Trends in Information Technology and Computer Science (ICRTITCS - 2012)*
- [6] Vishakha Patel, Gayatri Prabhu, Kiran Bhowmick, “Opinion Mining and Sentiment Analysis”, *International Journal of Computer Applications (0975 – 8887) Volume 131 – No.1, December 2015*
- [7] K.G. Nandha Kumar, T Christopher, “Opinion Mining”, *International Journal of Computer Applications (0975 – 8887) Volume 113 – No. 2, March 2015*
- [8] G. Vinodhini, RM. Chandrasekaran “Sentiment Analysis and Opinion Mining”, *International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 6, June 2012*
- [9] Wei Wei, “Analyzing Text Data for Opinion Mining”, Springer-Verlag Berlin Heidelberg 2011
- [10] Edison Marrese Taylor, Cristi’anRodr’íguez O., Juan D. Vel’asquez, Goldina Ghosh, and Soumya Banerjee, “Web Opinion Mining and Sentimental Analysis”, Springer-2012
- [11] Hsinchun Chen and David Zimbra, University of Arizona “AI and Opinion Mining” Part 1, *IEEE Intelligent Systems*, May/june 2010
- [12] Hsinchun Chen and David Zimbra, University of Arizona “AI and Opinion Mining” Part 2, *IEEE Intelligent Systems*, May/june 2010
- [13] SoujanyaPoria, Newton Howard, Amir Hussain, “Enhanced SenticNet with Affective Labels for Concept-Based Opinion Mining”, Published by the IEEE Computer Society, March/April 2013
- [14] Erik Cambria, Yunqing Xia, “New Avenues in Opinion Mining and Sentiment Analysis”, Published by the IEEE Computer Society, March/April 2013
- [15] Fuji Ren, Senior Member, IEEE, and Ye Wu, “Predicting User-Topic Opinions in Twitter with Social and Topical Context”, *IEEE Transactions On Affective Computing*, Vol. 4, No. 4, October-December 2013
- [16] ZhongwuZhai, Bing Liu, “Grouping for Opinion Mining”, Published by the IEEE Computer Society, July/August 2012
- [17] Lisette García-Moya, Henry Anaya-Sánchez, and Rafael, “ Retrieving Product Features and Opinions from Customer Reviews”, Published by the IEEE Computer Society, may/june 2013
- [18] Beema K S, Mitha Rachel Jose, “Co-Extracting Opinions from Online Reviews”, *International Journal of Computer Applications Technology and Research Volume 5– Issue 2, 95 - 98, 2016, ISSN:- 2319–8656*
- [19] Francis F. Balahadia and Benilda Eleonor V. Comendador, “ Adoption of Opinion Mining in the Faculty Performance Evaluation System”, *International Journal of Computer Theory and Engineering*, Vol. 8, No. 3, June 2016
- [20] I R Jayasekara and W M J I Wijayanayake, “Opinion Mining of Customer Reviews”, *International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.6, No.1, January 2016*

- [21] Mubarak Himmat, Naomie Salim, "Obstacles and Challenges of Opinion Mining of Customers Reviews", International Journal Of Enhanced Research In Management & Computer Applications, VOL. 2 ISSUE 2, FEB.-2013
- [22] HaseenaRahmath,Dr.Tanvir Ahmad, "Opinion Mining and Sentiment Analysis – A Review", IJSRD - International Journal for Scientific Research & Development| Vol. 2, Issue 03, 2014 | ISSN (online): 2321-0613
- [23] Maqbool Al-Maimani, NaomieSalim, Ahmed M. Al-Naamany, "Opinion Mining: Approaches, Resources And Challenges",Journal of Theoretical and Applied Information Technology, 20th May 2014. Vol. 63 No.2
- [24] Pang, B., Lee, L. and Vaithyanathan, S. "Thumbs up? Sentiment Classification Using Machine Learning Techniques." Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP'02), 2002.
- [25] Ganapathibhotla, G. and Liu, B. "Identifying Preferred Entities in Comparative Sentences." Proceedings of the 22nd International Conference on Computational Linguistics (COLING'08), 2008.
- [26] Turney, P. "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews." ACL'02, 2002.
- [27] Wei Huang, Xin Chen, Haibo Wang, "Product Information Retrieval Based on Opinion Mining", 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010),978-1-4244-5934-6/10/\$26.00 ©2010 IEEE
- [28] M. Hu and B. Liu, "Mining Opinion Features in Customer Reviews, " presented at 19th National Conf. on Artificial Intelligence, 2004.
- [29] L. B. Hu M, "Mining and summarizing customer reviews." presented at Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining 22:2525, 2004. \
- [30] Su-Ke Li, Zhi Guan, Li-Yong Tang and Zhong Chen "Exploiting Consumer Reviews for Product Feature Ranking" Journal of Computer Science and TechnologyJanuary Volume 27, Issue 3, pp 635-649 in 2012.
- [31] J. Han and M. Kamber, "Data Mining: Concepts and Techniques", third edition Publication Date: July 6, 2011.
- [32] W. Wang, H. Xu and W. Wan "Implicit feature identification via hybrid association rule mining", ESWA 8310 No. of Pages 14, Model 5G 3 January 2013.
- [33] A. Aamodt, H. A. Sandtorv, O. M. Winnem "Combining Case Based Reasoning and Data Mining - A way of revealing and reusing RAMS experience" Safety and Reliability; Proceedings of ESREL, Trondheim, June 16-19, 1998.
- [34] Y. Kim, S.-H. Myaeng, "Opinion Analysis based on Lexical Clues and their Expansion Information and Communications" Proceedings of NTCIR-6 Workshop Meeting, Tokyo, Japan in May 15-18, 2007
- [35] M. Baroni and S. Vegnaduzzo "Identifying subjective adjectives through web-based mutual information". In E. Buchberger, editor, In Proceedings of the Conference for the Processing of Natural Language and Speech (KONVENS), pages 17—24, in 2004.
- [36] Savary, L. and Zeitouni, K. "Indexed bit map (ibm) for mining frequent sequences", PKDD Knowledge Discovery in DatabasesPKDD, 9th European Conference on Principles and Practice of Knowledge Discovery in Databases', Vol. 3721 of Lecture Notes in Computer Science, Springer, Porto, Portugal, in 2005.
- [37] Harb, A., Plantié, M., Dray, G., Roche, M., Troussel, F. and Poncelet, P., "Web Opinion Mining: How to extract opinions from blogs?", International Conference on Soft Computing as Tran disciplinary Science in 2008.
- [38] V Balakrishnan, Mohammad R. Shakouri, H Hoodeh and Loo, Huck-Soo, "Predictions Using Data Mining and Case-based Reasoning: A Case Study for Retinopathy" international journal of computer and information engineering 6, 2012.
- [39] S.Poria, A. Gelbukh, E. Cambria, D.R. Das, S. Bandyopadhyay, ""Enriching SenticNet Polarity Scores Through Semi-Supervised Fuzzy Clustering" IEEE 12th International Conference on Data Mining Workshops in 2012.
- [40] Z. ZHAI, H. XU, J. LI and P. JIA "Sentiment Classification for Chinese Reviews Based on Key Substring Features" Natural Language Processing and Knowledge Engineering, 2009. NLPKE 2009. International Conference onDate of Conference: 24-27 Sept. 2009
- [41] T. Yao and L. Li "A Kernel-based Sentiment ClassificationApproach for Chinese Sentences" World Congress on ComputerScience and Information Engineering in March 31 2009-April 22009.
- [42] Khamparia, Aditya, Pandey, Babita. Knowledge and Intelligent computing methods in e-learning. *International Journal of Technology Enhanced Learning*, Vol. 7, 3, 2015:221-242.
- [43] Khamparia A, Pandey B. Architecture based comparison of semantic web service composition processes. *International Journal of Computer Applications*, Vol. 98-02, pp. 15-20, 2015
- [44] Khamparia A, Pandey B. A novel method of case representation and retrieval in CBR for E-learning. *Education and Information Technologies*, doi: 10.1007/s10639-015-9447-8 (2015):pp. 1-18.